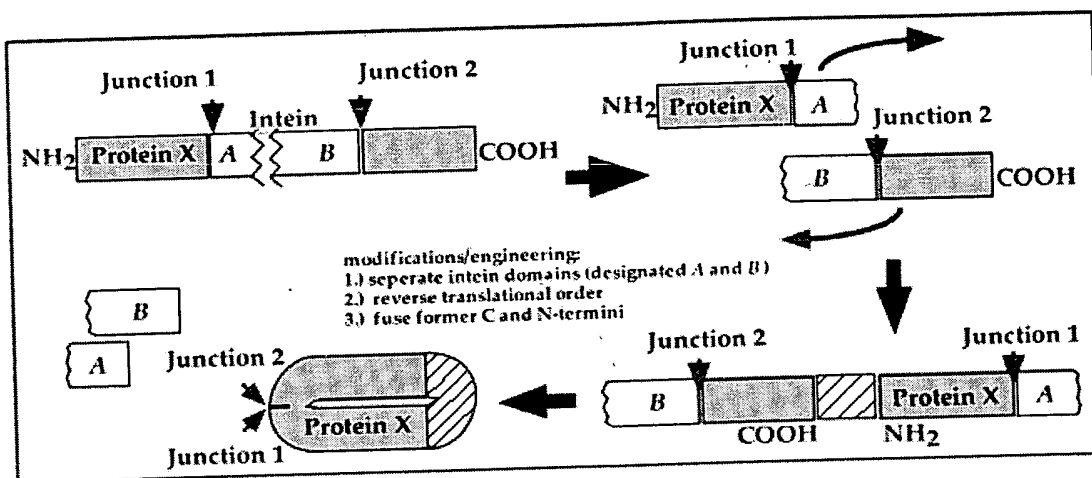
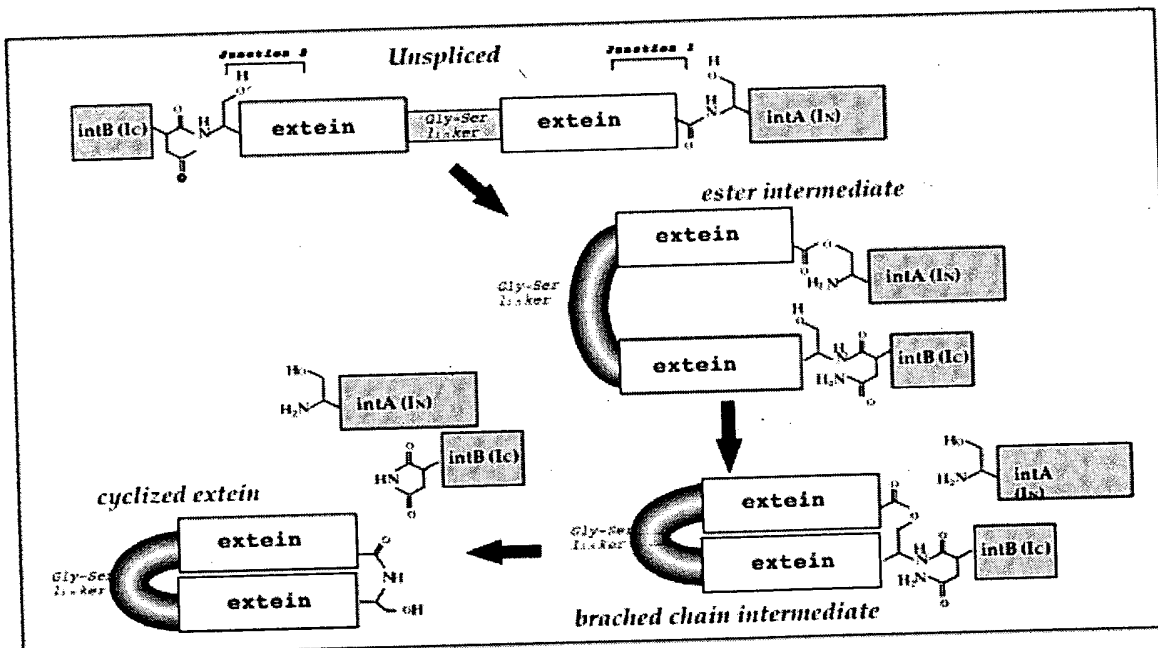


Figure 1

A



B



09400770-030601

FIGURE 3

D)

LCVAPETMILTEDGQFPIKDLEGKIIKVVWNGNEFSSVTVVKTGTEKELLELEVELSNGCT
LSCTPEHKFIIV
KSYTEAKKQKTDDNAIANAERVDAQDLKPRMKLIKFDLPTLFGNSEHDIKYPYTHGF
FCGDGTYTKYGKP
QLSLYGDKKELLTYLDVRTMTGLEDA SGRLNTWLPLDLAPKFDVPINSSLECRM EW
LAGYLDADGCVFRN
GTNESIQVSCIHLDFLKRIQLLLIGMGVTSKITKLHDEKITTMPDGKGGQKPYSCKPIW
RLFISSSGLYH
LSEQGFETRRLKWEPRQPQRNAERFVEVLKVNKTGRVDDTYCFTEPINHAGVFNGIL
TGQC

E)

GCFTKGTQVMMADGADKSIESIEVGDKVMGKDGMPREVVGLPRGYDDMYKVRQL
SSTRNAKSEGLMDFT
VSADHKLILKTKQDVKIA TRKIGGNTYTGVTFYVLEKTKTGIELVKAKTKVFGHHIH
GQNGAEKKAATFA
AGIDSKEYIDWIEARDYVQVDEIVKTSTTQMINPVHFESGKLGNWLHEHKQNKSLA
PQLGYLLGTWAGI
GNVKSSAFTMNSKDDVKLATRIMNYSSKLGMTCSSTESGELNVAENEEFFNNLGA
EKDEAGDFTFDEFT
DAMDELTINVHGAAASKKNLLWNALKSLGFRAKSTDIVKSIPQHIAVDDIVVRESLI
AGLVDAAGNVET
KSNGSIEAVVRTSFRHVARGLVKIAHSLGISSINIKDTHIDAAGVRQEFACIVNLTGA
PLAGVLSKCAL
ARNQTPVVKFTRDPVLFNFDLIKSAKENYYGITLAEETDHQFLLSNMALVHNC

F)

GCLSYATNQPYFLKSDNVNFSKLTSLKVS NHYILSATLELLIPFQYNRIYPIVSLIKREL
QTGYKV VYEL
DFYISVIVSTVEHYVLTNGWKRIELTVDDL VATLDIQYLIYNNTEVDLFSSNVIFSS
VINLICMNRIN
VYDFWIPKTNNFFVNALLVHNS

G)

GCISKFSHIMWSHVSKPLFNFSIKKSHMHNFNKNIYQLLDQGEAFISRQDKKTTYKIR
TNSEKYLELTSN
HKILTLRGWQRCDQLLCNDMITTQIGFELSRKKKYLLNCIPFSLCNFETLANINISNFQ
NVFDFAANPIP
NFIANNIIVHNS

FIGURE 3

H)

GCFAGTENVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQ
HRAHKSDSSREVPE
LLKFTCNATHEL VVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVS
KSYPISGPERA
NELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMOK
SKFHLTIEGPKVLA
YLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAYKDRKEPQVAKTVN
LYSKVVRGNGIRNN
LNTENPLWDAIVGLGFLKDGVKNIPLSTDNIGTRETFLAGLIDSDGYVTDEHGIKA
TIKTIHTSVRDG
LVSLARSLGLVVSNAEPAKVDMNGTKHKISYAIYMSGGDVLLNVLSKCAGSKKFR
PAPAAFARECRGF
YFELQELKEDDYGITLSDDSDHQFLLANQVVVHNC

I)

GCFAYGTRGALADGTTEKIGKIVNQKMDVEVMSYDPDQVVPKVVNWFNNGPA
EQFLQFTVEKSGGNG
KSQFAATPNHLIRTPAGWTEAGDLVAGDRVMAAEPHRLSDQQFQVVLGSLMGDGN
LSPNRRDRNGVRFRM
GHGAKQVDYLQWKTALLGNIKHSTHVNDKGATFVDFTPLPELAELQRAVYLGDK
KFLSEENFKALTPLA
LVFWYMDDGPFTVRSKGLQERTAGGSGRIEICVEAMSEGNRIRLRDYLDRDTHGLDV
RLRLSGAAGKSVLV
FSTASSAKFQELVAPYITPSMEYKLLPRFRGQGAVTPQFVEPTQRLV PARVLDVHVK
PHTRSMNRFDIEV
EGNHNYFVDGVMVHNS

J)

YCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDGRGEQEVLE
YELEDGSVIR
ATSDHRFLTDDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLDAGTIK

K)

KALALDTPLPTPTGWTAMGDVAVGDELLAVDEAPTRVVAATEVMLGRPCYEIEFSD
GTVIVADAQHQPWT
SYGIRTSACLRCGLDIIAAAGSTPRHAGRLTTAAAFMAPVLCIDSVRRVRSVPVRCVEV
DNAAHLYLARG
MVPTHNS

bioRxiv preprint doi: <https://doi.org/10.1101/022080>; this version posted October 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE 3

L)

GALAYDEPIYLSDGNIINIGEFVDKFFKKYKNSIKKEDNGFGWIDIGNENIYIKSFNKLS
LIIEDKRILR
VWRKKYSGKLIKITTKNRREITLTHDHPVYISKTGEVLEINAEMVKVGDYIYIPKNNTI
NLDEVIKVETV
DYNNGHIYDLTVEDNHTYIAGKNEGFAVSNC

M)

GALYDFSVIQLSNGRFVLIGDLVEELFKKYAEKIKTYKDLEYIELNEEDRFEVVSVP
DIKANKHVVS
RVWRRKVREGEKLIRIKTRTGNEIILTRNHPLFAFSNGDVVRKEAEKLVGDRVAVM
MRPPSPQTKA
VVDPAIYVKISDYLLVPNGKGMIVPNDGIPPEKAQYLLSVNSYPVKLVREVDEKLS
YLAGVILGDGY
ISSNGYYISATFDDEAYMDAFVSVVSDFIPNYVPSIRKNGDYTIVTVGSKIFAEMLSRI
FGIPRGRKS
MWDIPDVVLSNDDLMRYFIAGLFDADGYVDENGPSIVLVTKSETVARKIWYVLQRIG
IISTVSRVKSR
GFKEGELFRVIISGVEDLAKFAKFIPLRHSRKRAKLMEILRTKKPYRGRRTYRVPISSD
MIAPLRQML
GLTVAELSKLASYYAGEKVSESLIRHIEKGRVKEIRRSTLKGIALALQQIAKDVGNEE
AWVRAKRLQI
IAEGDVYWDEVVSVEEVDPKELGIEYVYDLTVEDDHNYVANGILVSNC

N)

PCVSGDTIVMTSGGPRTVAELEGKPFTALIRGSGYPCPSGFFRTCERDVYDLRTREGH
CLRLTHDHRVL
VMDGGLEWRAAGELERGDRLVMDDAAGEFPALATFRGLRGAGRQDVYDATVYGA
SAFTANGFIVHNC

O)

GCIDGKAKIIFENEGEEHLTTMEEMYERYKHLGEFYDEEYNRWGIDVSNVPIYVKSF
DPESKRVVKGKVN
VIWKYELGKDVTKEYEITNKGTKILTSPWHPFFVLTPDFKIVEKRADELKEGDILIGGM
PDGEDYKFIFD
YWLAGFIAGDGCDFKYHSHVKGHEYYIDRLRIYDYRIETFEIINDYLEKTFGRKYSIQ
KDRNIYYIDIKA
RNITSHYLLLEGIDNGIPPQILKEGKNAVLSFIAGLFDAEGHVSINKPGIELGMVKNKRL
IEDVTHYLNAL
GIKARIREKLRKDGIDYVLHVEEYSSLLRFYELIGKNLQNEEKREKLEKVL SNHKGGN
FGLPLNFNAFKE
WASEYGVFEKTNQSQTIAIINDERISLGQWHTRNRVSKAVLVKMLRKLYEATKDEEV

FIGURE 3

P)

NSILPEEWVPLIKNGKVKIFRIGDFVDGLMKANQGKVKKTGDTEVLEVAGIHAFSFD
RKSCKARVMAVKA
VIRHRYSGNVYRIVLNSGRKITITEGHSLFVYRNGDLVEATGEDVKIGDLLAVPRSVN
LPEKRERLNIVE
LLLNSPEETEDIILTIPVKGRKNFFKGMLRTLRLWIFGEEKRVRTASRYLRHLENLGYI
RLRKIGYDIID
KEGLEKYRTLIEKLVDVVRYNGNKREYLVEFNAVRDVISLMPEEELKEWRIGTRNG
FRMGTFVDIDEDFA
KLLGYYVSEGSARKWKNQTGGWSYTVRLYNENDEVLDDEMEHLAKKFFGKVVRGK
NYVEIPKKMAYIIFES
LCGTLAENKRVPEVIFTSSKGVRWAFLEGYFIGDGDVHPSKRVRLSTKSELLVNGLV
LLLNSLGVSAIKL
GYDSGVYRVYVNEELKFTEYRKKKNVYHSHIVPKDILKETFGKVVFQKNISYKKFREL
VENGKLDREKAKR
IEWLLNGDIVLDRVVEIKREYYDGYVYDLSVDEDENFLAGFGFLYAHNS

Q)

DSVTGETEIIKRNGKVEFVAIEELFQRVDYRIGEKEYCVLEGVEALTLDNRGRLVWK
SV
PYVMRHRTNKRIYRVWFTNSWYLDVTEHDHSLIGYMNTSKVKPGKPLKERLVEVKPG
ELGE
SVKSLITPNRAIAHGIRVNPIAVKLWELIGLLVGDGNWGGQSNWAKYNVGLSLGLDK
EEI
EEKILKPLKNTGIISNYYDKSKKGDVSILSKWLARFMVRYFKDESGSKRIPEFMFNLP
RE
YIEAFLRGLFSADGTVSLRKGVPVRLTSVNPELSSSVRKLLWLVGVSNSMFMVETNP
NRY
LGKESGTHSVHVRKDKHRFAERIGFLLDRKATKLSNLGGHTSKKRAYKYDFDLVY
PKK
VEEIAVDGYVYDIEVEGTHRFFANGILVHNT

R)

KCLLPEEKVVLPEIGLVTLRELFELANEVVVKDEEKEVRKLGKMLTGVDERGNVKL
LNALYVWRVAHK
GEMIRVKVNGWYSVTVTPEHPFLTNRGWVKAGELKEGDYIAIPRRVYGNEDIMKFS
KIAKELGIKGE
KEFYLAGASIDIPIKVLFLAPSKLVSAFLRGYFDAKGVVRENYIEVPLFEDLPLLILRFG
IVSRIEKS
TLKISGKRNLLEFRKHVGFTDSEKAKALDELISKAKESERYPIIEELRRLGLLFGFTRN

IPKECWGRGIRSF
 LRAYFDCNGGVKGNAIVLATASKEMSQEIAYALAGFGIISRIQEYRVIISGSDNVKKFL
 NEIGFINRNKL
 EKALKLVKKDDPGHDGLEINYELISYVKDRLRLSFFNDKRSWSYREAKEISWELMKE
 IYYRLDELEKLKE
 SLRGILIDWNEVAKRIEEVAEETGIRADELLEYIEGKRKLSFKDYIKIAKVLGIDVEHT
 IEAMRVFARK
 YSSYAEIGRRLGTWNSSVKTILESNAVNVEILERIRKIELELIEEILSDEKLKEGIAYLIF

FIGURE 3

U) cont.

LSQNELYWD
 EITKVEELRGEFIIYDLHVPGYHNFIAGNMPTVVHNT

V)

SCVTGDTKVYTPDEREVKIRDFMNYFENGLIKEVSNRIGRDTVIAAVSFNSRIVGHPV
 YRLTLESGRIIE
 ATGDHMF LTPEGWKQTYDIKEGSEVLVKPTLEGTPYEPDPRVIIDIKEFYNFLEKIERE
 HNLKPLKEAKT
 FRELITKDKEKILRRALELRAEIENGLTKREAEILELISADTWIPRAELEKKARISRTL
 NQILQRLEKK
 GYIERRIEGRKQFVRKIRNGKILRNAMDIKRILEEEFGIKISYTTVKKLLSGNVDGMAV
 RILKEVKEKWL
 VRYDDEKAGILARVVGFILGDGHLARNGRIFWNSKEELEMLANDLRKLGLKPSEIIE
 RDSSEIQGRKV
 KGRIYMLYVDNAAFHALLRFWKVEVGNKTKKGYTVPEWIKKGNLFVKREFLRGLF
 GADGTKPCGKRYNFN
 GIKLEIRAKKESLERTVEFLNDVADLLREFDVDSKITVSPTKEGFIIRLIVTPNDANYLN
 FLTRVGYAYA
 KDTYARLVGEYIRIKLAYKNIILPGIAEKAIELATVTNSTYAAKVLGVSRDFVVRNLK
 GTQIGITRDFMT
 FEEFMKERVNLNGYVIEKVIKKEKLG YLDVYDVT CARDHSFISNGLVSHNC

W)

NCLTSNSKILTDDGYIYKLEKLKEKLDLHIKIYNTEEKSSNILFVSERYADEKIIIRIK
 TESGRVLEGS
 KDHPVLTNLGYVPMGMLKEGDDVIVYPYEGVEYEEPSDEIILDEDDFAEYDKQIIKY
 LKDRGLLPLRMDN
 KNIGIARLLGFAGDGSIVKENGDRERLYVAFYGKRETLIKIREDLEKLGKASRIYSR
 KREVEIRNAY
 GDEYTSLCDNSIKITSKAFALFMHKLGMPIGKKTEQIYKIPEWIKKAPKWVKRNFLA
 GLFGADGSRVAF
 KNYTLPINLTMSKSEELKENILEFLNEIKLLLA EFDIESMIYEIKSLDGRVSYRLAIVG
 EESIKNFLGR
 INYEYSGEKKVIGLLAYEYLRRKDIAKEIRKKCIKRAKELYKKGVTVSEMLKMDEFR
 NEFISKRLIERAV

FOUO 0400360

X)
KCVDDGDTLVLTKEFGLIKIKELYEKLDGKGRKIVEGNEEWTELEKPITVYGYKDGKI
VEI
KATHVYKGVSSGMVEIRTRTGRKIKVTPIHRLFTGRVTKDGLILKEVMAMHVKPGD
RIAV
VKKIDGGEYIKLDSSNVGEIKVPEILNEELAEFLGYLMANGTLKSGIIEIYCDDSELLER
VNSLSLKLFGVGGRIVQKVDGKALVIQSKPLVDVLRRLGVPEDKKVENWKVPRELL
LSPS

X) cont.
 NVVRAFVNAYIKGKEEVEITLASEEGAYELSYLFAKLGIYVTISKSGEYYKVRVSRRG
 NL
 DTIPVEVNGMPKVLPHYEDFRKFAKSIGLEEVAENHLQHIIFDEVIDVRYIPEPQEVYDV
 T
 TETHNFVGGNMPTLLHNT

1043968

[illegible]

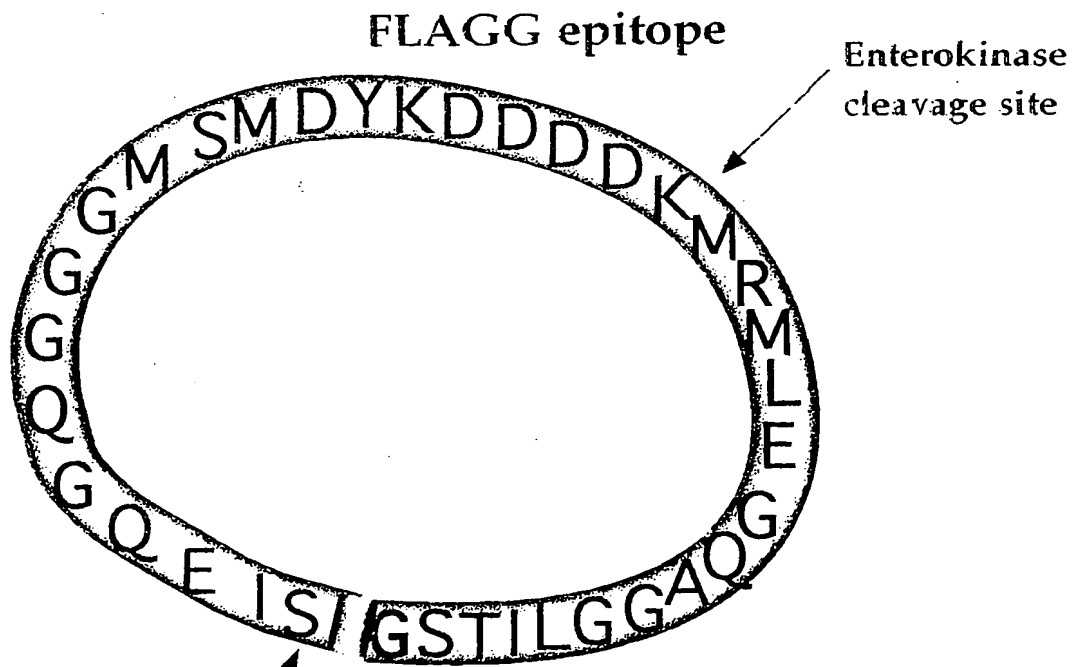
MESGSP EIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVAND

| | | | | |
|-------|---|-----------------------------------|---|-----|
| IIVHN | S | IEQGQGGGMSMDYKDDDDKMRMLEGQAGGLITS | G | CIS |
|-------|---|-----------------------------------|---|-----|

GDSLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESKVS RVFCT

GKKLVYILKTRLGRTIKATANHRFLTIDGWKRLDELSLKEHIALPRK

LESSSLQLSIHGYH



This is the only invariant extein -encoded amino acid
(depending on intein used this can be a cysteine, serine or threonine).

Figure 4B

CMV promoter →

| | | |
|---|-----------|----------|
| 1/1 | 31/11 | 61/21 |
| GCT TCG CGA TGT ACG GGC CAG ATA TAC GCG TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC A | | |
| 121/41 | 151/51 | 181/61 |
| TAC GGT AAA TGG CCC GCC TGG CTG ACC GCC CAA CGA CCC CCG CCC ATT GAC GTC AAT AAT GAC GTA T | | |
| 241/81 | 271/91 | 301/101 |
| TTT ACG GTA AAC TGC CCA CTT GGC AGT ACA TCA AGT GTA TCA TAT GCC AAG TAC GCC CCC TAT TGA C | | |
| 361/121 | 391/131 | 421/141 |
| GGA CTT TCC TAC TTG GCA GTA CAT CTA CGT ATT AGT CAT CGC TAT TAC CAT GGT GAT GCG GTT TTG G | | |
| 481/161 | 511/171 | 541/181 |
| CCA CCC CAT TGA CGT CAA TGG GAG TTT GTT TTG GCA CCA AAA TCA ACG GGA CTT TCC AAA ATG TCG T | | |
| 601/201 | 631/211 | 661/221 |
| CTA TAT AAG CAG AGC TCT CTG GCT AAC TAG AGA ACC CAC TGC TTA CTG GCT TAT CGA AAT TAA TAC G | | |
| 721/241 | 751/251 | 781/261 |
| CTg tcg acT GGA GGA ACC ATG GAG TCC GGA tca cca gaa ata gaa aag ttg tct cag agt gat att t | intB (lc) | |
| M E S G S P E I E K L S Q S D I Y | | |
| 841/281 | 871/291 | 901/301 |
| ttg act gtg cca gga cca cat aac ttt gtc gcc aat gac atc att gtc cat aac agt ATC GAA CAA g | | |
| L T V P G P H N F V A N D I I V H N S I E O G | | |
| 961/321 | 991/331 | 1021/341 |
| ATG ctg gag ggc caa gca ggt gga CTG ATC ACC agt ggc TGC ATC AGT GGA GAT AGt ttg atc agc t | | |
| M L E G Q A G G L I T S G C I S G D S L I S L | | |
| 1081/361 | 1111/371 | 1141/381 |
| ttt gaa ata tgg gca att aat gaa cag acg atg aag cta gaa tca gct aaa gtt agt cgt gta ttt t | | |
| F E I W A I N E O T M K L E S A K V S R V F C | | |
| 1201/401 | 1231/411 | 1261/421 |
| aag gca aca gca aat cat aga ttt tta act att gat ggt tgg aaa aga tta gat gag cta tct tta a | | |
| K A T A N H R F L T I D G W K R L D E L S L K | | |
| 1321/441 | 1351/451 | 1381/461 |
| GAT cca tgg tta cca TGA caa ttg GCG GCC GCT CGA GTC TAG AGG GCC CGC GGT TCG AAG GTA AGC C | | |
| D P W L P * | | |
| 1441/481 | | |
| ATC ACC ATT GAG TTT AAA CCC GCT GAT | | |

F09020-020066

FIGURE 5

A)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGACG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AaAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT

FIGURE 5

E)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
 GACTCCATCGTTTCTATTACGG
 AGAC
 TGGAGTCGAAGAGGTTTTTGGATTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
 AATGACATCATTGTCCATAAC
 AGTG
 AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
 CCATCGGCGACGGCCCCGTGCT
 GCTG
 CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
 AAGCGCGATCACATGGTCCTGC
 TGGA
 GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
 GAACGGGGAATTCTCGCAGGTA
 GACA
 AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
 TCGAGCTGGACGGCGACGTAAA
 CGGC
 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
 ACCCTGAAGTTCATCTGCACCA
 CCGG
 CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGAG
 TGCTTCAGCCGCTACCCCGAC
 CACA
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGA
 CGGC
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
 ATCGAGCTGAAGGGCATCGACT
 TCAA
 GGAGGACGGCAACATCCTGGGGGCACAAGCTGGAGTACAACAGCCACAA
 CGTCTATATCATGGCCGACAAG
 CAGA
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
 AAAAGCTGATATGCATCTCCGG
 AaAT
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
 ATGAAAAAGATTTTGAAATAT
 GGGC
 AgTTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTGT
 ACTGGCAAAAAGCTAGTTTAT
 ATTT
 TAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA
 CTATTGATGGTTGGAAAAGATT
 AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

F)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIK
ISGN
SLISLASTGKRVS IKDLLDEKDFEIWAVNEQTMKLESAKVS RVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

G)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTGgCcGTGCCAGGGCCCCATAACTTTGTGGCCA
ATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

1090070-03001

TGCTTCAGCCGCTACCCCGAC
 CACA
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGA
 CGGC
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
 ATCGAGCTGAAGGGCATCGACT
 TCAA
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA
 CGTCTATATCATGGCCGACAAG
 CAGA
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
 AAAAGCTGATATGCATCTCCGG
 AAAT
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
 ATGAAAAAGATTTTGAAATAT
 GGGC
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTGT
 ACTGGCAAAAAGCTAGTTTAT
 ATTT
 TAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
 CTATTGATGGTTGGAAAAGATT
 AGAT
 GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
 TACAATTAGGCCTCCGCGGCC
 AGTA
 CCCCTACGACGTCCCGGACTACGCTATCGATTAA

H)
 MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLAVPGPHNFVANDIIVHNSEEDLGSS
 VQLADHYQQNTPIGDG
 PVLL
 PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
 VSKGEELFTGVVPILVELDG
 DVNG
 HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ
 HDFFKSAMPEGYVQERTIFF
 KDDG
 NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGI
 KVNFKIRHNIEDLEQKLC
 ISGN
 SLISLASTGKRVSIGDLLDEKDFEIWAINETMKLESASVSRVFCTGKKLVYILKTRL
 GRTIKATANHRFLTIDGW
 KRLD
 ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

D)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCTGTACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTG
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AaAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT

TGCTTCAGCCGCTACCCCGAC
 CACA
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGA
 CGGC
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
 ATCGAGCTGAAGGGCATCGACT
 TCAA
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA
 CGTCTATATCATGGCCGACAAG
 CAGA
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
 AAAAGCTGATATGCATCTCCGG
 AaAT
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
 ATGAAAAAGATTTTGAAATAT
 GGGC
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
 ACTGGCAAAAAGCTAGTTTAT
 ATTT
 TAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA
 CTATTGATGGTTGGAAAAGATT
 AGAT
 GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
 TACAATTAGGCCTCCGCGGCC
 AGTA
 CCCCTACGACGTCCCGGACTACGCTATCGATTAA

L)
 MESGSPEIEKLSQSDIYWDSIVSITETGVVEEVDLTVPGPHNFVANDIIVHNSEEDLGSS
 VQLADHYQQNTPIGDG
 PVLL
 PDNHYLSTQSALS KDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
 VSKGEELFTGVVPILVELDG
 DVNG
 HKFSVSGEGEGDATYGKLT LKFICTTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKQ
 H DFFKSAMPEGYVQERTIFF
 KDDG
 NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGI
 KVNFKIRHNIEDLEQKLIC
 ISGN
 SLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESAKVS RVFCTGKRLVYILKTRL
 GRTIKATANHRFLTIDGW
 KRLD
 ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

M)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
 GACTCCATCGTTTCTATTACGG
 AGAC
 TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
 AATGACATCATTGTCCATAAC
 AGTG
 AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
 CCATCGGCGACGGCCCCGTGCT
 GCTG
 CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
 AAGCGCGATCACATGGTCCTGC
 TGGA
 GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
 GAACGGGGAATTCTCGCAGGTA
 GACA
 AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
 TCGAGCTGGACGGCGACGTAAA
 CGGC
 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
 ACCCTGAAGTTCATCTGCACCA
 CCGG
 CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGAG
 TGCTTCAGCCGCTACCCCGAC
 CACA
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGA
 CGGC
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGC
 ATCGAGCTGAAGGGCATCGACT
 TCAA
 GGAGGACGGCAACATCCTGGGGGCACAAGCTGGAGTACAACAGCCACAA
 CGTCTATATCATGGCCGACAAG
 CAGA
 AGAACGGCATCAAGGTGAAGTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
 AAAAGCTGATATGCATCTCCGG
 AGAT
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
 ATGAAAAAGATTTTGAAATAT
 GGGC
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
 ACTGGCAAAAAGCTAGTTTAT
 ATTT
 TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAaATTTTAAAC
 TATTGATGGTTGGAAAAGATT
 AGAT •

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

N)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIK
ISGD
SLISLASTGKRVS IKDLLDEKDFEIWAINEQTMKLESAKVS RVFCTGKKLVYILKTRL
GRTIKATANHKFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

O)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTG CAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCC GTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTG CAG

TGCTTCAGCCGCTACCCCGAC
 CACA
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGA
 CGGC
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
 ATCGAGCTGAAGGGCATCGACT
 TCAA
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA
 CGTCTATATCATGGCCGACAAG
 CAGA
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
 AAAAGCTGATATGCATCTCCGG
 AGAT
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
 ATGAAAAAGATTTTGAAATAT
 GGGC
 AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTGT
 ACTGGCAAAAAGCTAGTTTAT
 ATTT
 TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA
 CTATTGATGGTTGGAAAAGATT
 AGAT
 GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
 TACAATTAGGCCTCCGCGGCC
 AGTA
 CCCCTACGACGTCCCGGACTACGCTATCGATTAA

P)
 MESGSPEIEKLSQSDIYWDSIVPITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
 VQLADHYQQNTPIGDG
 PVLL
 PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
 VSKGEELFTGVVPILVELDG
 DVNG
 HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ
 HDEFFKSAMPEGYVQERTIFF
 KDDG
 NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGI
 KVNFKIRHNIEDLEQKLIC
 ISGD
 SLISLASTGKRVSIGDLLDEKDFEIWAINEQTMKLESAKVS RVFCTGKKLVYILKTRL
 GRTIKATANHRFLTIDGW
 KRLD
 ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

Q)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
 GACTCCATCGTTTCTATTACGG
 AGAC
 TGGAGTCGAAGAGGTTTTTGATTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
 AATGACATCATTGTCCATAAC
 AGTG
 AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
 CCATCGGCGACGGCCCCGTGCT
 GCTG
 CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
 AAGCGGATCACATGGTCCTGC
 TGGA
 GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
 GAACGGGGAATTCTCGCAGGTA
 GACA
 AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGG
 TCGAGCTGGACGGCGACGTAAA
 CGGC
 CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
 ACCCTGAAGTTCATCTGCACCA
 CCGG
 CAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
 TGCTTCAGCCGCTACCCCGAC
 CACA
 TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGA
 CGGC
 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGC
 ATCGAGCTGAAGGGCATCGACT
 TCAA
 GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAA
 CGTCTATATCATGGCCGACAAG
 CAGA
 AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
 AAAAGCTGATATGCATCTCCGG
 AGAT
 AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
 ATGAAAAAGATTTTGAAATAT
 GGGC
 AGTTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
 ACTGGCAAAAAGCTAGTTTAT
 ATTT
 TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTAA
 CTATTGATGGTTGGAAAAGATT
 AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

R)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTCLKFICTTGKLPVPWPTLVTTLTYGVCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGD
SLISLASTGKRVS IKDLLDEKDFEIWAVNEQTMKLESAKVS RVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

02203889

Figure 6

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|--|--|---|--|--|--|--|--|--|--|--|--|--|--|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| 1/1 | | | | | | | | | | | | 31/11 | | | | | | | | | | | | 61/21 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| GCT TCG CGA TGT ACG GGC CAG ATA TAC GCG | | | | | | | | | | | | TTG ACA TTG ATT ATT GAC TAG TTA TTA ATA GTA ATC A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 121/41 | | | | | | | | | | | | 151/51 | | | | | | | | | | | | 181/61 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TAC GGT AAA TGG CCC GCC TGG CTG ACC GCC | | | | | | | | | | | | CAA CGA CCC CCG CCC ATT GAC GTC AAT AAT GAC GTA T | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 241/81 | | | | | | | | | | | | 271/91 | | | | | | | | | | | | 301/101 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TTT ACG GTA AAC TGC CCA CTT GGC AGT ACA | | | | | | | | | | | | TCA AGT GTA TCA TAT GCC AAG TAC GCC CCC TAT TGA C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 361/121 | | | | | | | | | | | | 391/131 | | | | | | | | | | | | 421/141 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| GGA CTT TCC TAC TTG GCA GTA CAT CTA CGT | | | | | | | | | | | | ATT AGT CAT CGC TAT TAC CAT GGT GAT GCG GTT TTG G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 481/161 | | | | | | | | | | | | 511/171 | | | | | | | | | | | | 541/181 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CCA CCC CAT TGA CGT CAA TGG GAG TTT GTT | | | | | | | | | | | | TTG GCA CCA AAA TCA ACG GGA CTT TCC AAA ATG TCG T | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 601/201 | | | | | | | | | | | | 631/211 | | | | | | | | | | | | 661/221 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CTA TAT AAG CAG AGC TCT CTG GCT AAC TAG | | | | | | | | | | | | AGA ACC CAC TGC TTA CTG GCT TAT CGA AAT TAA TAC G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 721/241 | | | | | | | | | | | | 751/251 | | | | | | | | | | | | intB (lc) 781/261 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CTg tcg acT GGA GGA ACC | | | | | | | | | | | | ATG GAG TCC GGA | | | | | | | | | | | | tca cca gaa ata gaa aag ttg tct cag agt gat att t | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | M E S G | | | | | | | | | | | | S P E I E K L S O S D I Y | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 841/281 | | | | | | | | | | | | 871/291 | | | | | | | | | | | | 901/301 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ttg act gtg cca gga cca cat aac ttt gtc | | | | | | | | | | | | gcc aat gac atc att gtc cat aac | | | | | | | | | | | | agt ATC GAA CAA g | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| L T V P G P H N F V | | | | | | | | | | | | A N D I I V H N | | | | | | | | | | | | S I E O G | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 961/321 | | | | | | | | | | | | 991/331 | | | | | | | | | | | | 1021/341 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ATG ctc gag ggc caa gca ggt gga CTG ATC | | | | | | | | | | | | ACC agt | | | | | | | | | | | | ggc TGC ATC AGT GGA GAT AGt ttg atc agc t | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| M L E G Q A G G L I T S | | | | | | | | | | | | G | | | | | | | | | | | | C I S G D S L I S L | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1081/361 | | | | | | | | | | | | 1111/371 | | | | | | | | | | | | 1141/381 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ttt gaa ata tgg gca att aat gaa cag acg | | | | | | | | | | | | atg aag cta gaa tca gct aaa gtt agt cgt gta ttt t | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| F E I W A I N E O T M K L E S A K V S R V F C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1201/401 | | | | | | | | | | | | 1231/411 | | | | | | | | | | | | 1261/421 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aag gca aca gca aat cat aga ttt tta act | | | | | | | | | | | | att gat ggt tgg aaa aga tta gat gag cta tct tta a | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| K A T A N H R F L T I D G W K R L D E L S L K | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| GAT | | | | | | | | | | | | ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| D | | | | | | | | | | | | M V S K G E E L F T G V V P I L V E L D G D | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Y G K L T L K F I C T T G K L P V P W P T L V | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| K O H D F F K S A M P E G Y V O E R T I F F K | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| L V N R I E L K G I D F K E D G N I L G H K L | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| N G I K V N F K I R H N I E D G S V Q L A D H | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CAC TAC CTG AGC ACC CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG C | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| H Y L S T O S A L S K D P N E K R D H M V L L | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

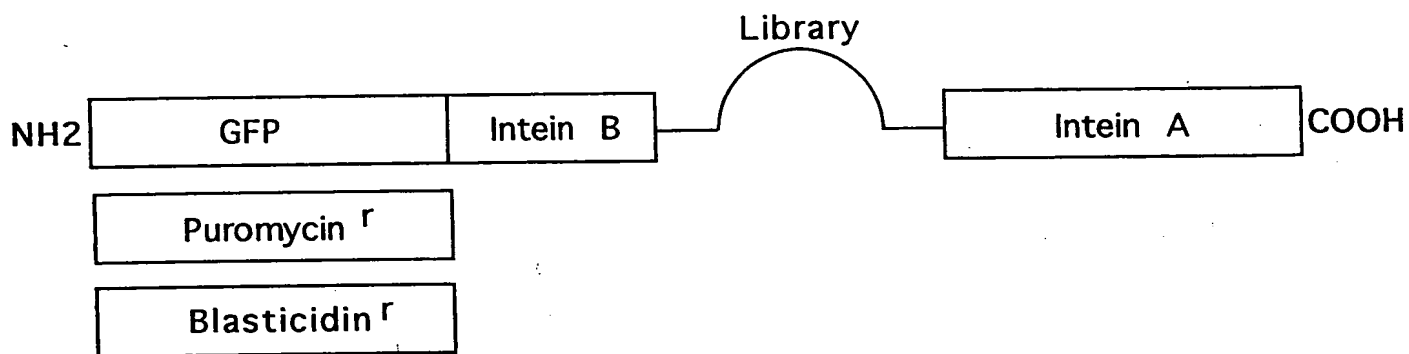
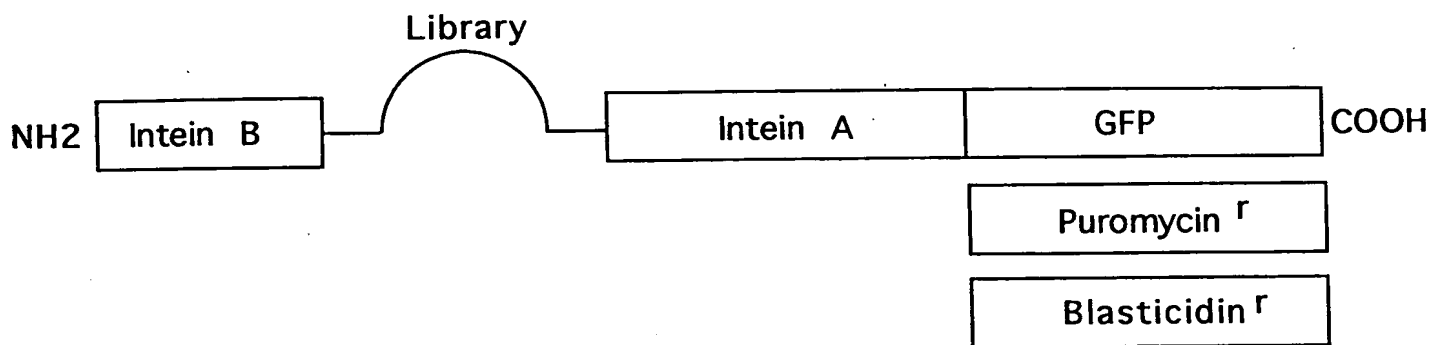


Figure 7

00000770-020804

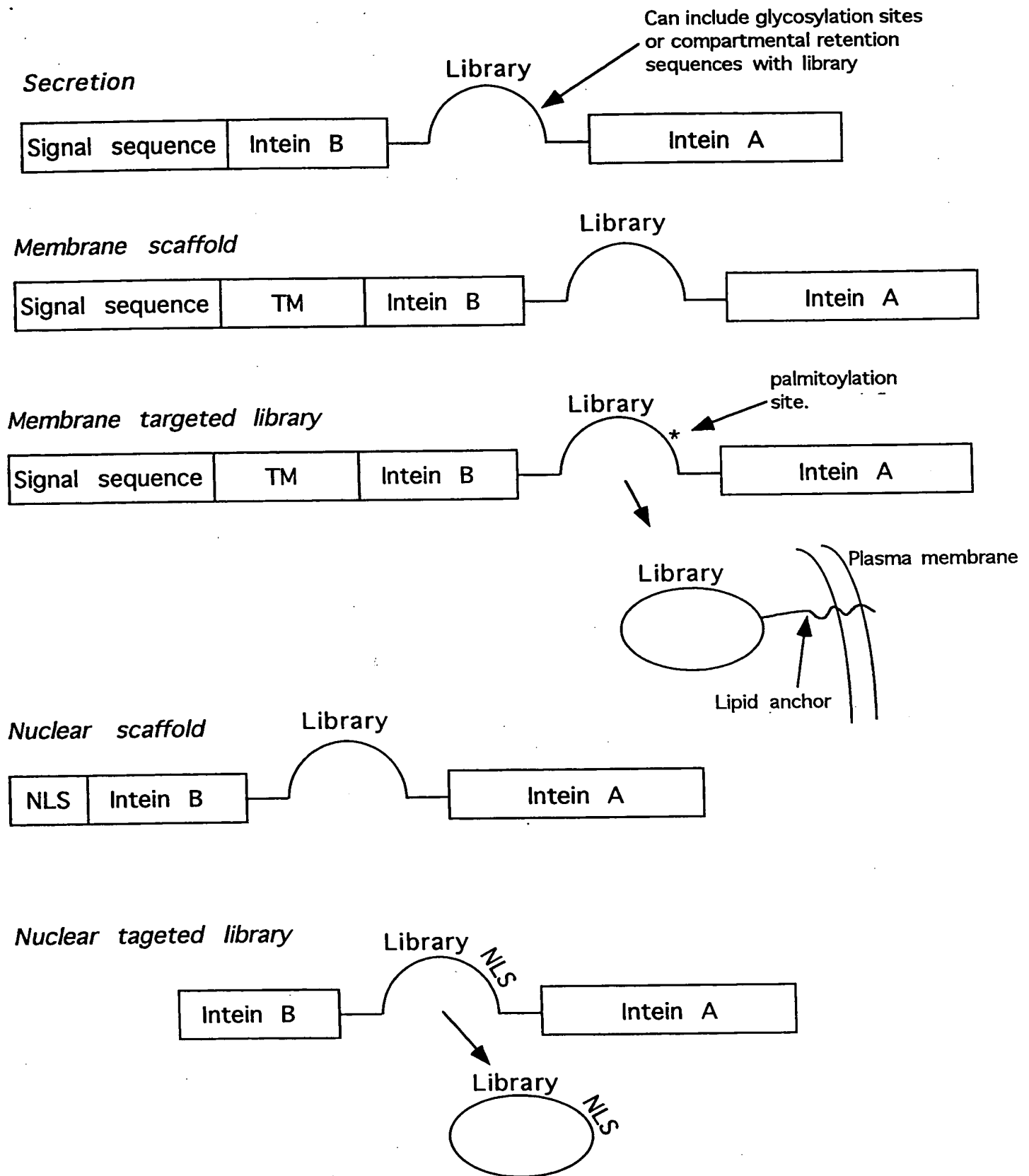


Figure 8

Membrane-tethered
scaffold

1. Use a PCR mutagenesis or shuffling approach to mutate intein domains
2. Create a retroviral library of mutants
3. Infect cells and screen for those most efficient at cyclization (assayed indirectly by monitoring the release of IntA-TetRVP16 from its membrane location)

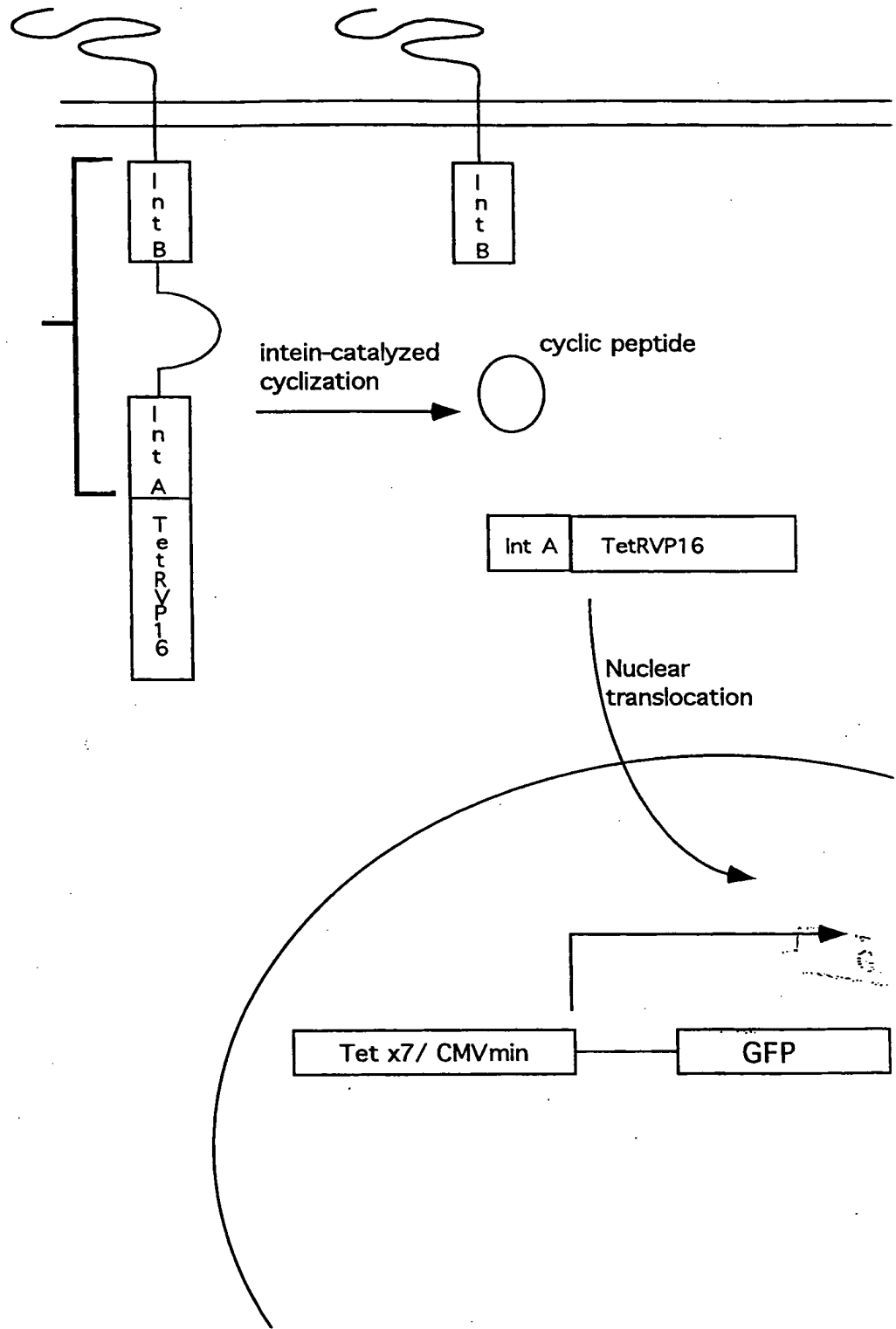
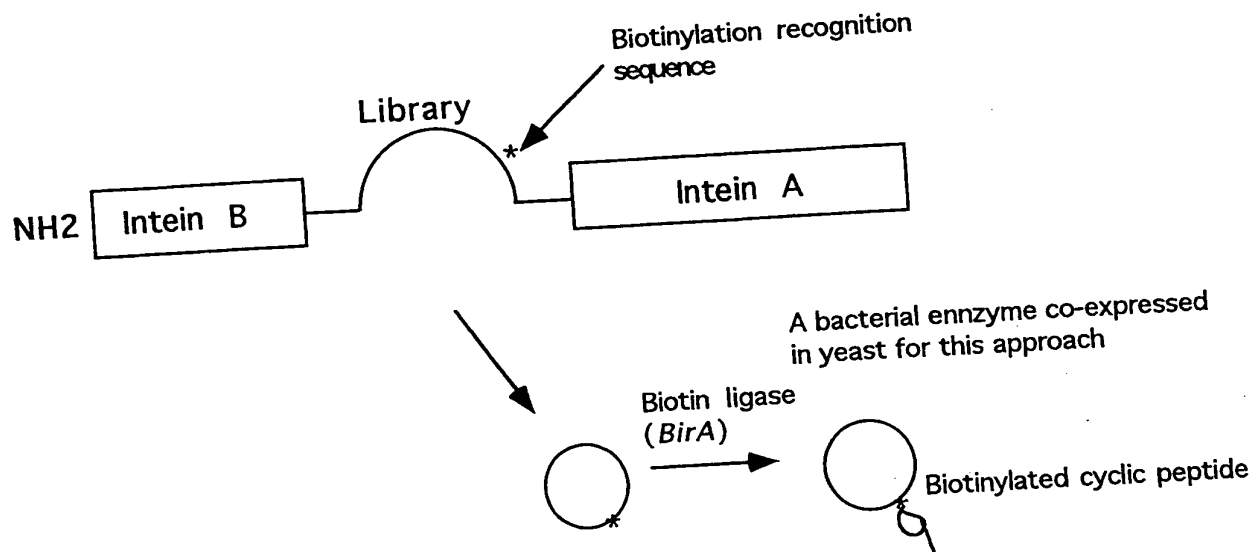


Figure 9



typical cDNA target/transactivation domain fusion utilized in yeast two hybrid systems

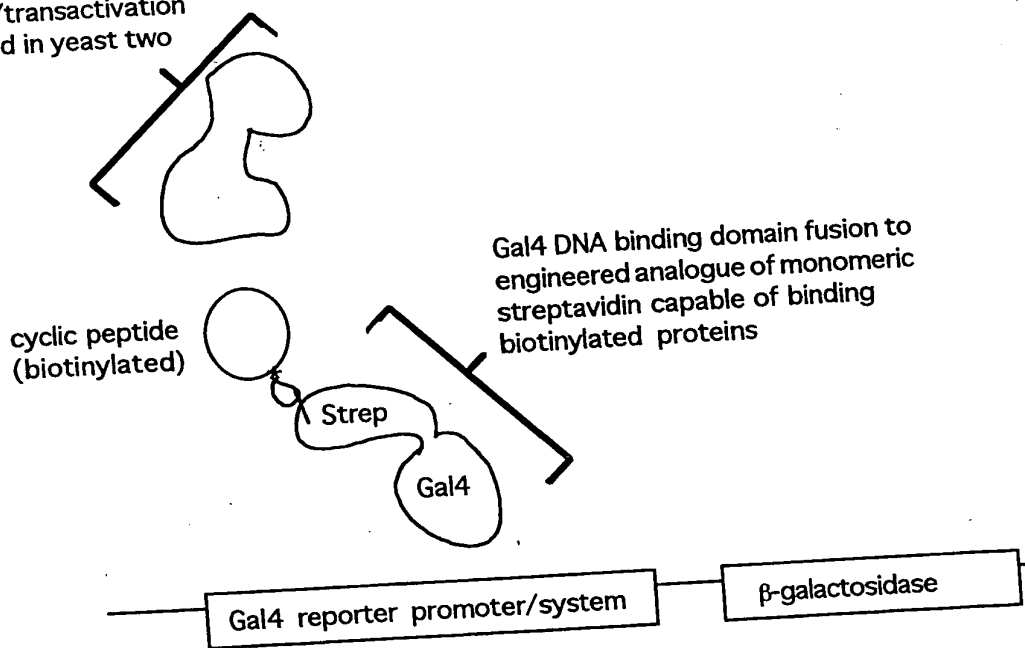
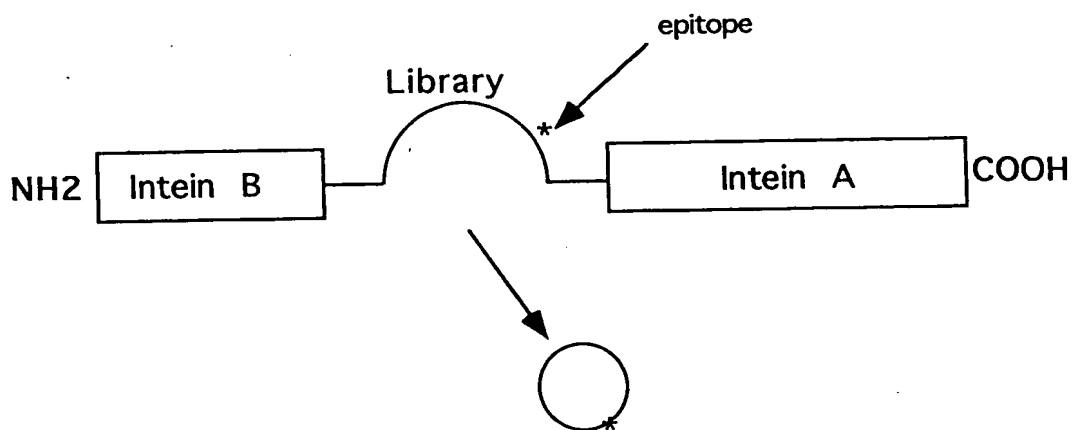
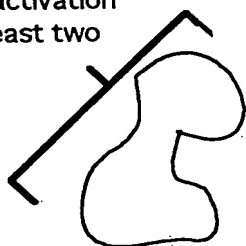


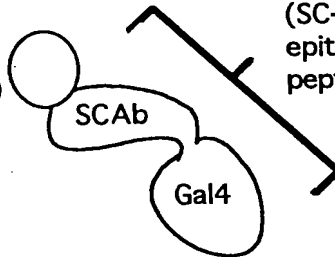
Figure 10



typical cDNA target/transactivation domain fusion utilized in yeast two hybrid systems



cyclic peptide (with epitope)



Gal4 DNA binding domain fusion to engineered single chain antibody (SC-Ab) capable of binding to the epitope present within the cyclized peptide

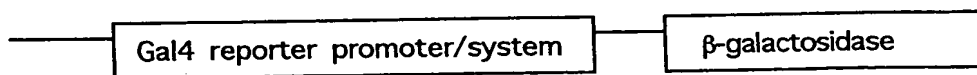


Figure 11

५

m... ..
 int h (i) {
 if (i < 0) return 0;
 if (i < 10) return i;
 if (i < 20) return i * 2;
 if (i < 30) return i * 3;
 if (i < 40) return i * 4;
 if (i < 50) return i * 5;
 if (i < 60) return i * 6;
 if (i < 70) return i * 7;
 if (i < 80) return i * 8;
 if (i < 90) return i * 9;
 if (i < 100) return i * 10;
 return i * 11;
 }

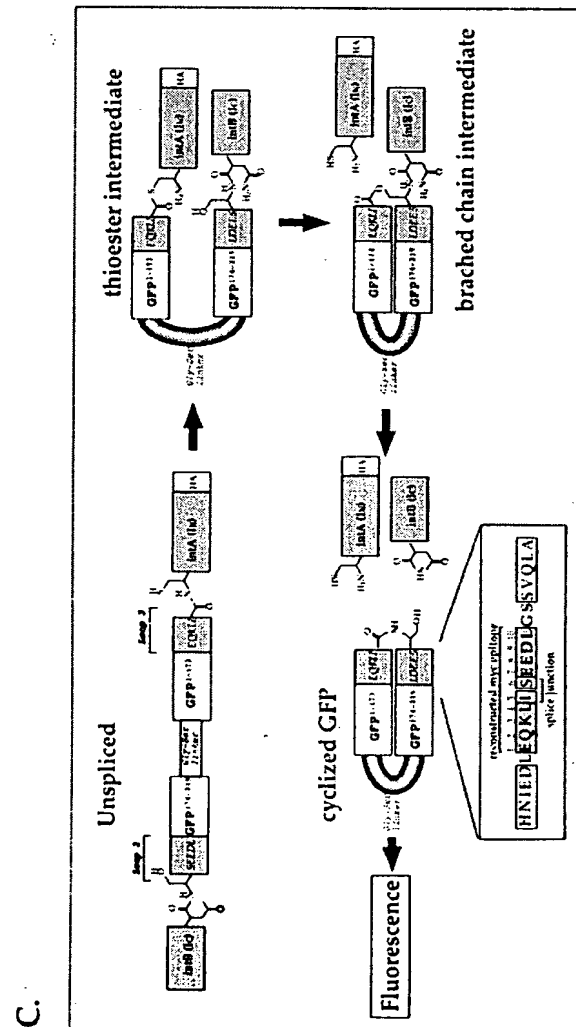
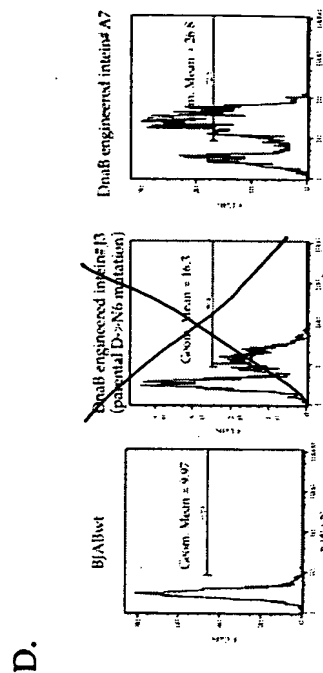
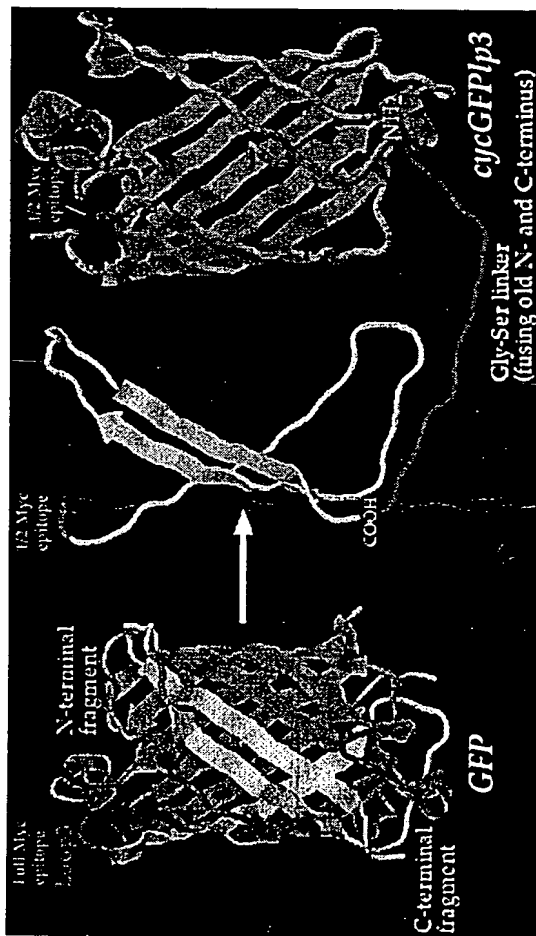
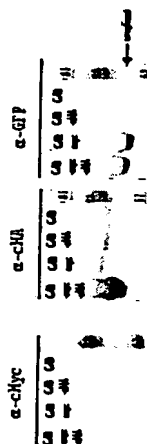
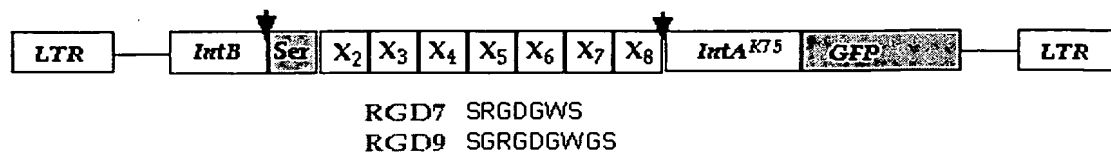


Figure 12

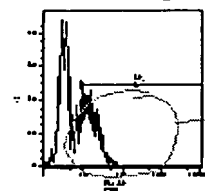


Retroviral library construct: control RGD inserts



Infect cell line

sort GFP⁺ population



establish cell line

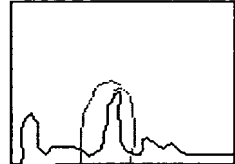
LTR

Lysate

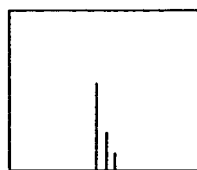
10kDa MWCO

10kDa lysate

collect fraction



MALDI-TOF
expected mass



C18

RP-HPLC

cyclic
standard

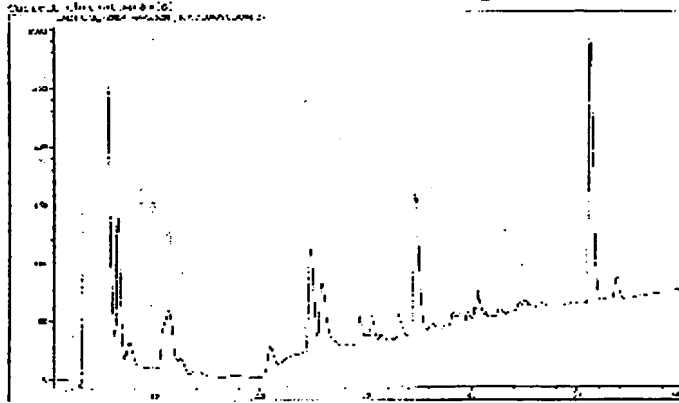
C18

retention
time
standard



Figure 15A

A5T4-RGD7 lysate



A5T4-RGD7 lysate (HPLC 34-35min. fraction standard)

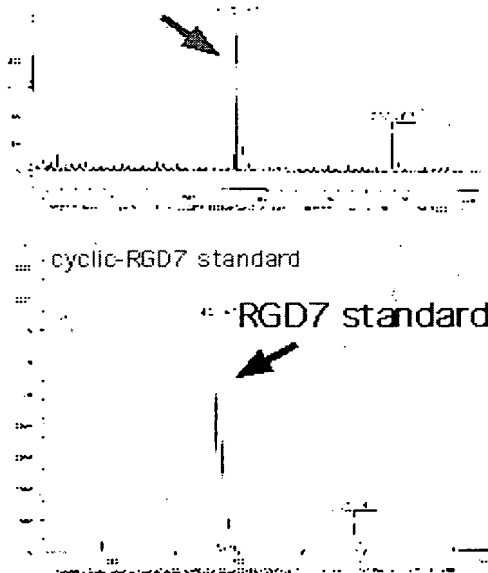
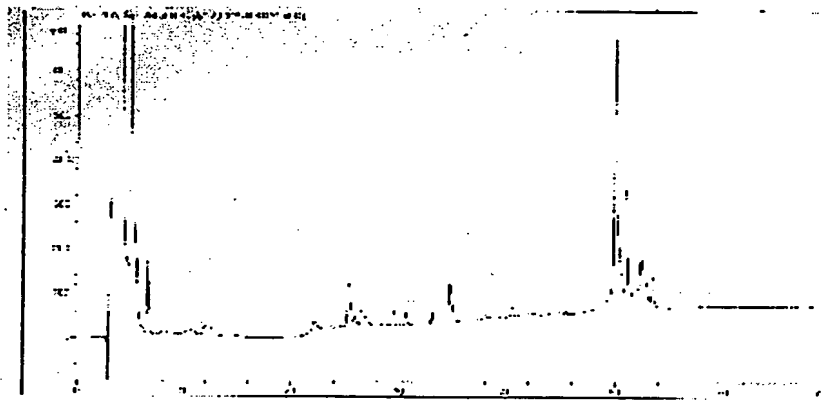


Figure 15 B

A5T4-RGD9 lysate



RP-HPLC fraction
(33-34 min.)

A5T4-RGD9 lysate (HPLC 33-34 min.
fraction standard (expect: 860.4)

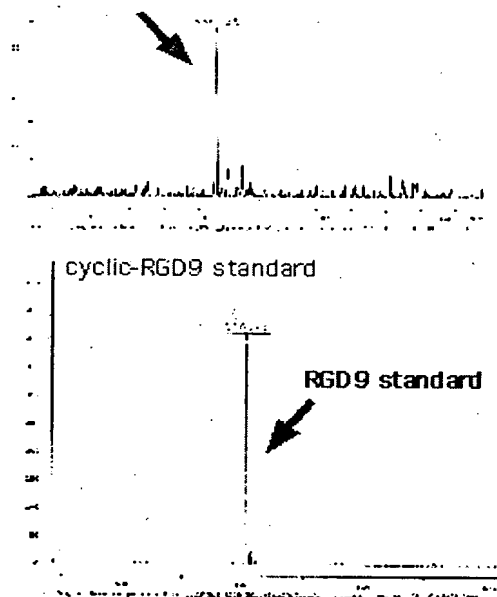


Figure 15C

LC/MS fragmentation fingerprinting

01/28/01 05:57:43 PM

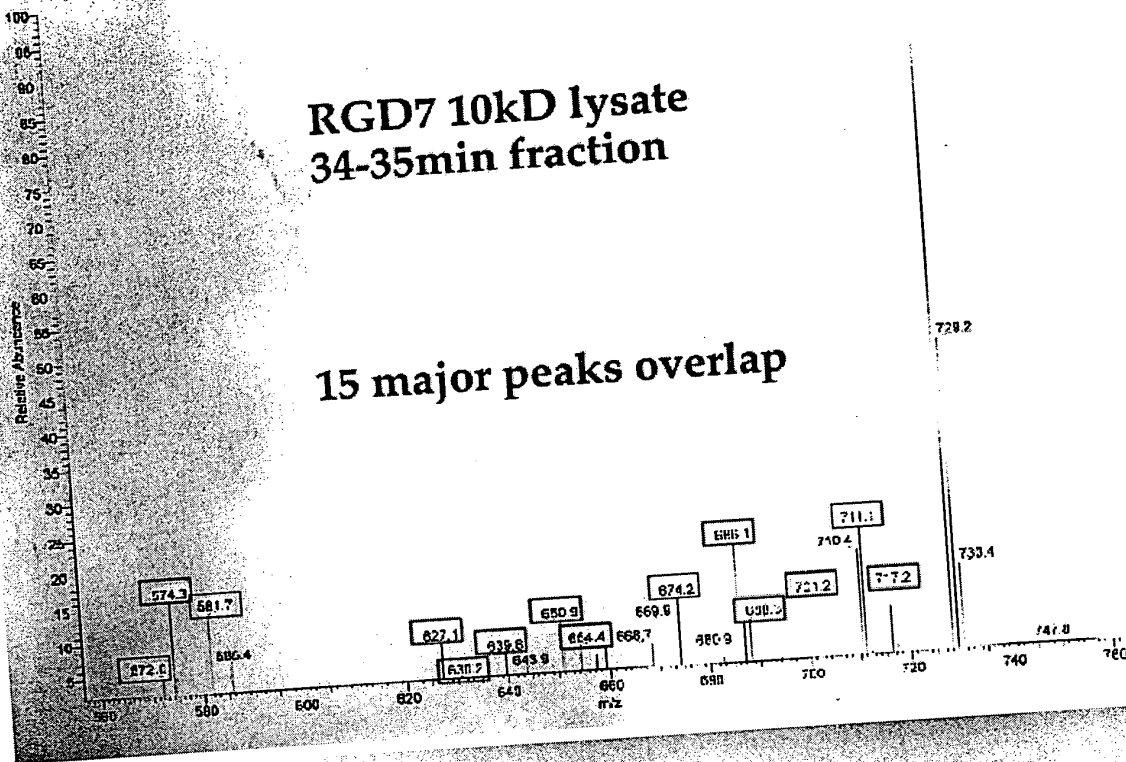
D:\LCQ Data\blank_star_RGD7

Blank_star_RTD: #1766 RT: 38.50 AV: 1 NL: 4.95E4
T: + C d Full m/z 748.46 @ 12.00 (185.03-1535.03)

728.2

RGD7 10kD lysate
34-35min fraction

15 major peaks overlap



D:\LCQ Data\RGD7_Std

RGD7_Std #417 RT: 4.70 AV: 1 NL: 1.50E5
T: + C d Full m/z 742.70 @ 12.00 (185.03-1535.03)

01/28/01 04:35:03 PM

RGD7 standard

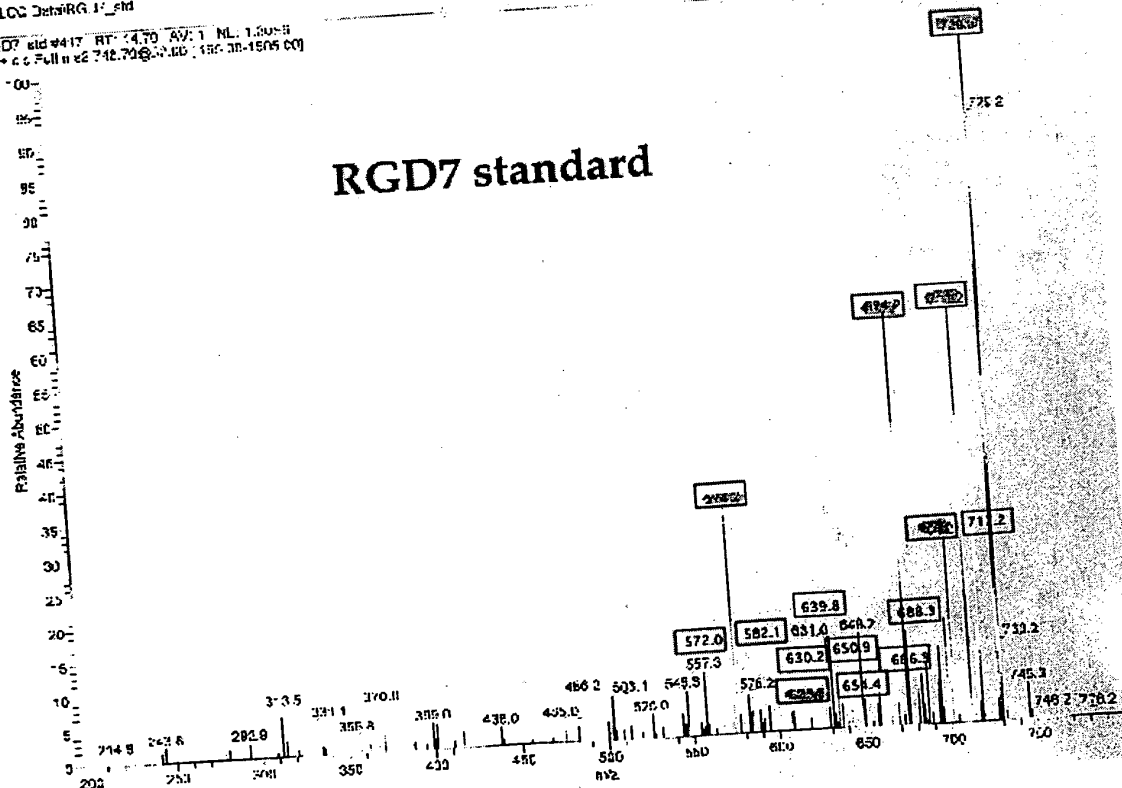


Figure 15D

17-336